Aix-Marseille Université **M1 MAS**
Statistique exploratoire

Cours d'Analyse en Composantes Principales (ACP)

adapté d'un polycopié rédigé par Marie-Christine Roubaud (I2M)

Historique : L'ACP est une des plus anciennes méthodes factorielles. Elle a été conçue par Karl Pearson (1901) et intégrée à la statistique par Harold Hotelling (1933).

Contexte d'application : On observe sur n individus p variables **quantitatives** X^1, X^2, ..., X^p présentant des relations multiples que l'on veut analyser.
L'ACP est une méthode de description et de résumé d'un tel tableau de données.

Objectifs principaux :

- Donner la "meilleure" représentation plane de l'ensemble des individus, minimisant les déformations du nuage des points.
- Donner une représentation graphique plane explicitant "au mieux" les liaisons des variables initiales.
- Résumer *l'information liée à l'inertie*, contenue dans le tableau initial (n,p) dans un tableau de plus faible dimension (n,k), $k < p$, (réduction de la dimension) en la détruisant la moins possible. Ce tableau résumé est obtenu en remplaçant les variables initiales X^j, $j = 1, \ldots, p$ par un petit nombre de variables non corrélées, C^j, $j = 1, \ldots, k$ combinaisons linéaires des X^j et résumant "au mieux" l'information initiale.

Limites de l'ACP :

- Ne permet pas le traitement de variables **qualitatives**.
- Ne permet de détecter que d'éventuelles liaisons linéaires entre variables.
- Extrait des données uniquement l'information liée au critère d'inertie qui n'est pas forcement le critère le plus adapté pour caractériser les données

L'ACP présente de nombreuses variantes selon les transformations apportées au tableau de données. Parmi ces variantes, l'ACP sur un tableau où les colonnes sont centrées et réduites, appelée **ACP normée**, est la plus fréquemment utilisée.

Remarque sur les notations :

Plusieurs dimensions apparaissent en ACP. Ainsi dans la mesure du possible, on notera les différents objets avec des indices de la façon suivante :
$objet_i$ est un objet lié à un individu i,
$objet^j$ est un objet lié à une variable j,
$objet_k$ est un objet lié à un $k$ième axe de l'ACP et appartenant à l'espace des individus $\mathbb{R}^p$,
$objet^k$ est un objet lié à un $k$ième axe de l'ACP et appartenant à l'espace des variables $\mathbb{R}^n$.

1 Définitions

1.1 Espace des individus et espace de variables

On observe p variables quantitatives X^1, X^2, ..., X^p sur n individus.
Les observations ainsi obtenues constituent une matrice X ayant n lignes (**individus**) et p colonnes (**variables**) :

$$X = \begin{pmatrix} x_1^1 & x_1^2 & \dots & x_1^p \\ x_2^1 & x_2^2 & \dots & x_2^p \\ \vdots & \vdots & \vdots & \vdots \\ x_n^1 & x_n^2 & \dots & x_n^p \end{pmatrix}$$

où x_i^j est la valeur prise par la j ème variable sur le $i$ème individu.

Pour les n individus :
On associe à chaque individu i un point x_i de $\mathbb{R}^p$ de coordonnées :

$$x_i = {}^t(x_i^1, x_i^2, \cdots, x_i^p) \qquad \text{(transposée de la ième ligne de la matrice)}.$$

Chaque individu peut-être alors représenté dans $\mathbb{R}^p$ appelé **espace des individus**.

Pour les p variables :
A chaque variable X^j est associé le vecteur x^j correspondant à la $j$ème colonne de X :

$$x^j = \begin{pmatrix} x_1^j \\ x_2^j \\ \vdots \\ x_n^j \end{pmatrix}$$

Chaque variable X^j peut-être alors représentée par un vecteur de $\mathbb{R}^n$ appelé **espace des variables**.
Remarque : par la suite par abus de langage, on pourra nommer le vecteur x_i "l'individu i", et le vecteur x_j "la variable j". Ainsi une variable est assimilée à la liste des n valeurs qu'elle prend sur les n individus, et un individu est assimilé à la liste des p valeurs que lui affectent les p variables.

Définition 1. *On affecte à chaque individu un* **poids** p_i *reflétant son importance par rapport aux autres individus avec* $p_i > 0$ *et* $\sum_{i=1}^n p_i = 1$.
On appelle **matrice des poids** *la matrice diagonale* (n, n) *dont les éléments diagonaux sont les poids* p_i. *Elle sera notée* $D = diag(p_1, p_2, \cdots, p_n)$.

Exemple : On peut considérer que tous les individus ont la même importance : $p_i = \frac{1}{n}$, pour tout $i = 1, \cdots, n$.

1.2 Nuage des individus

Définition 2. *On appelle* **nuage des individus**, *l'ensemble des points* x_i *munis de leurs poids :*

$$\mathcal{M}\{(x_i, p_i)\,;\, i = 1, \dots, n\}\,.$$

Rappels :
- Moyenne empirique de la variable x^j : $\bar{x}^j = \sum_{i=1}^n p_i x_i^j$.
- Variance empirique de x^j : $s_j^2 = \sum_{i=1}^n p_i (x_i^j - \bar{x}^j)^2 = \sum_{i=1}^n p_i (x_i^j)^2 - (\bar{x}^j)^2$.
- Covariance empirique de x^j et x^k : $s_{jk} = \sum_{i=1}^n p_i (x_i^j - \bar{x}^j)(x_i^k - \bar{x}^k) = \sum_{i=1}^n p_i x_i^j x_i^k - \bar{x}^j \bar{x}^k$.
- Corrélation empirique de x^j et x^k : $r_{jk} = \frac{s_{jk}}{s_j s_k}$.

Définition 3. *Le point* g *de* $\mathbb{R}^p$ *dont les coordonnées sont* $g = {}^t(\bar{x}^1, \bar{x}^2, \cdots, \bar{x}^p)$ *est le centre de gravité (***le barycentre***) du nuage de points* $\mathcal{M}$.

Pour ramener l'origine du repère au barycentre des individus (i.e centrer le nuage autour de son barycentre), on centre les variables. A chaque variable observée x^j on associe sa variable centrée y^j définie par $\forall i \in \{1,\dots,n\},\ y_i^j = x_i^j - \bar{x}^j$. Ainsi :

$$y^j = x^j - \begin{pmatrix} \bar{x}^j \\ \bar{x}^j \\ \vdots \\ \bar{x}^j \end{pmatrix}$$

On obtient alors le tableau des données centrées Y :

$$Y = \left[y^1, y^2, \cdots, y^p\right] = \begin{pmatrix} y_1^1 & \dots & y_1^j & \dots & y_1^p \\ y_2^1 & \dots & y_2^j & \dots & y_2^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ y_n^1 & \dots & y_n^j & \dots & y_n^p \end{pmatrix}.$$

Définition 4. *On appelle* **nuage centré des individus**, *l'ensemble des points de* $\mathbb{R}^p$

$$\mathcal{N} = \{(y_i, p_i)\,;\, i = 1,\dots,n\},$$

où pour tout $1 \le i \le n$, $y_i = {}^t(x_i^1 - \bar{x}^1, x_i^2 - \bar{x}^2, \cdots, x_i^p - \bar{x}^p)$ *(c'est à dire la transposée de la ième ligne de la matrice* Y*).*

Par la suite on sera amenés aussi à étudier les variables centrées et réduites, notées $z^1,\dots,z^p$, et définies par $\forall i \in \{1,\dots,n\},\ z_i^j = \frac{x_i^j - \bar{x}^j}{s_j}$. On obtient alors le tableau des données centrées et réduites Z :

$$Z = \left[z^1, z^2, \cdots, z^p\right] = \begin{pmatrix} z_1^1 & \dots & z_1^j & \dots & z_1^p \\ z_2^1 & \dots & z_2^j & \dots & z_2^p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_n^1 & \dots & z_n^j & \dots & z_n^p \end{pmatrix}.$$

Soit V la matrice de covariance et R la matrice de corrélation, qui joueront un rôle important par la suite :

$$V = \begin{pmatrix} s_1^2 & s_{12} & \dots & s_{1p} \\ s_{21} & s_2^2 & \dots & s_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ s_{p1} & s_{p2} & \dots & s_p^2 \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix}.$$

Remarques :

1) La matrice R est la matrice de variance-covariance des données centrées et réduites et résume la structure des dépendances linéaires entre les p variables.

2) Les matrices V et R sont carrées de taille p, symétriques et semi-définies positives : pour tout $u \in \mathbb{R}^p$, ${}^tuVu \ge 0$ et ${}^tuRu \ge 0$. D'après le théorème spectral, elles admettent donc p (en comptant les multiplicités) valeurs propres réelles positives ou nulles.

Notons aussi $D_{1/S}$ la matrice diagonale des inverses des écarts-types :

$$D_{1/S} = diag(1/s_1, \cdots, 1/s_p).$$

Proposition 1. *1) On a* $Z = YD_{1/S}$.

2) On peut montrer que la matrice de variance-covariance du tableau X *peut s'écrire sous la forme*

$$V = {}^tXDX - g{}^tg = {}^tYDY.$$

3) On a $R = D_{1/S}VD_{1/S} = {}^tZDZ$.

1.3 Distance dans l'espace des individus $\mathbb{R}^p$

Une question primordiale : **comment mesurer la distance entre individus ?** Le problème doit être résolu avant toute étude statistique car les résultats obtenus en dépendent dans une large mesure. En statistique chaque dimension correspond à un caractère qui s'exprime avec une unité particulière. La question est par exemple de calculer la distance entre 2 individus décrits par 3 caractères tels que *l'âge*, *le salaire* et *le nombre d'enfants*. Si on veut donner des importances différentes à chaque caractère, on peut prendre pour distance

$$d(x_1, x_2) = \sqrt{m_1(x_1^1 - x_2^1)^2 + m_2(x_1^2 - x_2^2)^2 + \cdots + m_p(x_1^p - x_2^p)^2}\,,$$

ce qui revient à multiplier par $\sqrt{m_j}$ chaque variable X^j. Cependant, cette formule sous-entend que les axes sont orthogonaux (formule de Pythagore), mais en statistique c'est par pure convention que l'on représente les variables par des axes orthogonaux, on aurait pu prendre des axes obliques. Ainsi la distance entre deux individus x_1 et x_2 peut être définie de manière générale par la forme quadratique :

$$d_M(x_1, x_2) = \sqrt{{}^t(x_1 - x_2)M(x_1 - x_2)} = ||x_1 - x_2||_M\,,$$

où M est une matrice symétrique définie positive : pour tout $u \in \mathbb{R}^p$, ${}^tuMu \geq 0$ et si ${}^tuMu = 0$ alors $u = 0_{\mathbb{R}^p}$. La matrice M admet p valeurs propres réelles strictement positives.

Cette métrique est associée au produit scalaire sur $\mathbb{R}^p$ suivant :

$$< u, v >_M = {}^tuMv.$$

En pratique, on utilise le plus souvent l'une des métriques suivantes :

1) $M = I_d$, la distance est la distance euclidienne usuelle, on parle d'**ACP canonique ou simple**. **Utilisation :** lorsque les variables sont de même unité de mesure et de **même ordre de grandeur**.

2) $M = D_{1/S^2}$, où D_{1/S^2} est la matrice diagonale des inverses des variances définie par $D_{1/S^2} = D_{1/S}D_{1/S}$. Le choix de cette métrique est (pratiquement*) équivalent à diviser chaque variable (colonne) par son écart-type et à en faire une ACP canonique. On parle alors d'**ACP normée**. Ici la distance ne dépend plus des unités de mesure puisque x_i^j/s_j est un grandeur sans dimension. Cette métrique donne à chaque variable la même importance quelle que soit sa dispersion. **Utilisation :** lorsque les variables ne sont pas de même unité de mesure ou ne sont pas de **même ordre de grandeur**.

***Remarque :** En toute rigueur, il y a de légères différences entre faire une ACP des données centrées avec la métrique $M = D_{1/S^2}$ et faire une ACP des données centrées et réduites avec la métrique $M = I_d$. Par la suite c'est cette 2ème ACP qui sera appelée **ACP normée**.

Précision : Comme indiqué, utiliser les variables centrées et réduites permet de rendre les grandeurs comparables, quelles que soient les unités de mesures des variables initiales. Ainsi par exemple, si la distribution des données est relativement symétrique autour de la tendance centrale :

$z_i \simeq 0$ signifie que l'individu a une valeur x_i moyenne,
$|z_i| \simeq 1$: valeur de x_i assez habituelle,
$z_i \geq 2$: valeur x_i particulièrement élevée,
$z_i \leq -2$: valeur x_i particulièrement faible.

1.4 Distance dans l'espace des variables $\mathbb{R}^n$

Chaque variable est représentée par un vecteur x^j de $\mathbb{R}^n$. Il apparaît naturel ici de considérer que deux variables sont proches si elles varient dans le même sens (ex : Taille et poids).

Pour formaliser cette notion, il faut munir $\mathbb{R}^n$ d'une métrique i.e choisir une matrice M (n,n) symétrique et définie positive. Ici, il n'y a pas d'hésitation, on choisit $M = D$, la matrice diagonale des poids (dite **métrique des poids**) pour les raisons suivantes.

Supposons que les données sont centrées (variables centrées) et soit $y^1, y^2, \ldots, y^p$ les variables centrées associées à $x^1, x^2, \ldots, x^p$.

Dans l'espace des variables muni de la métrique des poids, on remarque que le produit scalaire et la norme s'expriment comme la covariance et la variance, i.e

$< y^j, y^k >_D = s_{jk}$,

$||y^j||_D^2 = s_j^2$.

De plus l'angle θ_{jk} entre les vecteurs y^j et y^k est donné par

$$\cos\theta_{jk} = \frac{< y^j, y^k >_D}{||y^j||_D||y^k||_D} = \frac{s_{jk}}{s_j s_k} = r_{jk} .$$

En résumé, lorsque les variables sont centrées et représentées par des vecteurs de $\mathbb{R}^n$ muni de la métrique des poids :

- La longueur du vecteur correspond à l'écart-type de la variable associée ,
- Le cosinus de l'angle de deux vecteurs représente la corrélation linéaire des deux variables associées.

Remarque : Dans l'espace des individus, on s'intéresse aux distances entre les points individus alors que dans l'espace des variables, on s'intéressera plutôt aux angles en raison de la popriété précédente. Ainsi par convention, dans les graphiques, les individus sont représentés par des points et les variables par des vecteurs.

1.5 Inertie d'un nuage de points

Définition 5. *On appelle* **inertie totale du nuage des individus**, *notée I, la moyenne pondérée des carrés des distances des points au centre de gravité :*

$$I = \sum_{i=1}^{n} p_i d_M^2(x_i, g) = \sum_{i=1}^{n} p_i ||x_i - g||_M^2 .$$

L'inertie mesure la dispersion des points individus autour du centre de gravité g, elle est parfois appelée **variance du nuage**.

Exemple : Soit $\mathcal{M}$ le nuage des cinqs points suivants de $\mathbb{R}^2$:

$$A(1,0) \quad B(-1,0) \quad C(0,0), \quad D(-1,1), \quad E(0,2) ,$$

à qui on attribue le même poids.

1) On considère la distance euclidienne usuelle sur $\mathcal{M}$. Faire un graphique représentant ces points dans un repère orthonormé et calculer l'inertie du nuage.
2) *Changement de métrique :* déterminer l'expression analytique de la distance d_M sur $\mathbb{R}^2$ associé à la matrice $M = D_{1/S^2}$. Calculer l'inertie dans ce cas.

Propriété : On peut montrer que

$$I = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} p_i p_j ||x_i - x_j||_M^2 ,$$

c'est à dire l'inertie correspond à la moyenne des carrés de toutes les distances entre les individus.

Remarque : L'inertie du nuage $\mathcal{M}$ est évidemment égale à l'inertie du nuage centré $\mathcal{N}$

$$I = \sum_{i=1}^{n} p_i ||x_i - g||_M^2 = \sum_{i=1}^{n} p_i ||y_i||_M^2 = \sum_{i=1}^{n} p_i d_M^2(y_i, O)\,.$$

Proposition 2. *Expression matricielle de l'inertie. On a* $I = tr(VM)$.

Conséquence :
- Si $M = I_d$, alors $I = \sum_{i=1}^{n} s_j^2$, *somme des variances empiriques des p variables*
- Si $M = D_{1/S^2}$, alors $I = p$.
  Dans ce cas l'inertie est égale au nombre de variables et ne dépend pas de leurs valeurs.

Nous travaillerons dans toute la suite avec les données centrées.

Définition 6. *On appelle* **inertie du nuage des individus $\mathcal{N}$ expliquée (portée) par le sous-espace vectoriel** F *de* $\mathbb{R}^p$*, l'inertie du nuage projeté sur F, c'est à dire*

$$I_F(\mathcal{N}) = \sum_{i=1}^{n} p_i d_M^2(\hat{y}_i^F, O) = \sum_{i=1}^{n} p_i ||\hat{y}_i^F||_M^2\,,$$

où $\hat{y}_i^F$ *= projection orthogonale de* y_i *sur* F.

Remarque : On a évidemment le résultat suivant pour tout sous-espace vectoriel F : $I_F(\mathcal{N}) = I(\hat{\mathcal{N}}^F)$.

Inertie expliquée (ou portée) par un axe :
Soit u un vecteur M-normé i.e $||u||_M = 1$, Δ_u la droite vectorielle engendrée par u et $\hat{y}_i^u$ la projection orthogonale de y_i sur Δ_u.
Pour tout i variant de 1 à n on a :

$$\hat{y}_i^u = < y_i\,,\, u >_M \; u$$

L'inertie expliquée par Δ_u est donnée par

$$I_{\Delta_u} = I(\hat{\mathcal{N}}^u) = \sum_{i=1}^{n} p_i ||\hat{y}_i^u||_M^2 = {}^t u M V M u.$$

Propriété : Si $\mathbb{R}^p = F \oplus F^{\perp}$, on peut montrer qu'on a la décomposition suivante (faire un dessin en dimension 2) :

$$I = I_F + I_{F^{\perp}}\,.$$

Donc la quantité $I_{F^{\perp}}$ peut-être considérée comme **la déformation du nuage projeté sur** F :

$$I_{F^{\perp}} = \sum_{i=1}^{n} p_i ||y_i - \hat{y}_i^F||_M^2.$$

L'inertie totale se décompose donc pour tout F de $\mathbb{R}^p$ comme la somme de
- l'inertie du nuage projeté sur F, $I(\hat{\mathcal{N}}^F)$,
- la déformation du nuage $\mathcal{N}$ par projection orthogonale sur F, $I_{F^{\perp}}$.

Décomposition de l'inertie expliquée : De la même manière on montre que si $F = F_1 \oplus F_2$ et $F_1 \perp F_2$ alors

$$I_F = I_{F_1} + I_{F_2}$$

c'est à dire Inertie expliquée par F= Inertie expliquée par F_1 + Inertie expliquée par F_2.

2 Formulation du problème de l'ACP et solution

Rappelons que l'objectif principal est d'obtenir une représentation fidèle du nuage des individus de $\mathbb{R}^p$ en le projetant sur un espace de faible dimension. Le choix de l'espace de projection s'effectue selon le critère de l'inertie : on cherche le sous-espace de dimension k portant l'inertie maximale du nuage. Ceci revient à déformer le moins possible les distances en projection.

2.1 Première étape : centrage de données

On translate le centre de gravité g de $\mathcal{M}$ à l'origine O de $\mathbb{R}^p$.

2.2 Formulation mathématique du problème

Le problème peut s'écrire sous la forme suivante :
Trouver le sev E_k de dimension k ($k < p$), tel que l'inertie expliquée par E_k soit **maximale**.

Définition 7. *On appelle* **sous-espace principal** *de dimension k, tout sev de dimension k répondant à la question.*

2.3 Résolution du problème

D'après la décomposition de l'inertie sur deux sev M-orthogonaux, on montre que les sous-espaces principaux E_k (les solutions) sont emboîtés :

Théorème 1. *Soit E_k le sous espace vectoriel de dimension $k < p$ portant l'inertie maximale du nuage, alors le sous-espace de dimension $k+1$ portant l'inertie maximale est*

$$E_k \oplus \Delta_{u_{k+1}}$$

où u_{k+1} est un vecteur M-ortogonal à E_k et $\Delta_{u_{k+1}}$ est la droite vectorielle portant l'inertie maximale parmi toute les droites M-orthogonale à E_k .

On suppose dans ce qui suit que les vecteurs u_j sont M-normés.

2.3.1 Procédure

- Rechercher un axe Δ_{u_1} maximisant l'inertie expliquée $I_{\Delta_{u_1}} = {}^t u_1 MVMu_1$. On note $E_1 = \Delta_{u_1}$.
- Rechercher un axe Δ_{u_2} orthogonal à E_1, maximisant l'inertie expliquée $I_{\Delta_{u_2}} = {}^t u_2 MVMu_2$. On note $E_2 = E_1 \oplus \Delta_{u_2}$.
- ...
- Rechercher un axe Δ_{u_k} orthogonal à E_{k-1} maximisant l'inertie expliquée $I_{\Delta_{u_k}} = {}^t u_k MVMu_k$. On note $E_k = E_{k-1} \oplus \Delta_{u_k}$.

2.3.2 Recherche des vecteurs principaux

Conséquence du théorème spectral : Toute matrice $A(p,p)$ réelle M-symétrique admet p valeurs propres réelles et une base de vecteurs propres M-orthonormés.

Théorème 2. *Soit r le rang de V (qui est aussi le rang de Y). Soit une famille M-orthonormée $(u_1, u_2, \cdots, u_r)$ formée par les vecteurs propres de la matrice VM rangés par ordre décroissant des valeurs propres non nulles $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$. Alors u_1, u_2, ..., u_r engendrent respectivement les axes recherchés, appelés* **axes principaux d'inertie** *de l'ACP.*

Remarque : D'après la Proposition 2 on a $I = tr(VM) = \sum_{j=1}^{r} \lambda_j$. Ainsi (voir plus loin) $I_{E_r} = I$ et $I_{E_r^\perp} = 0$, c'est pourquoi on arrête l'algorithme à $k = r$.

Définition 8. *Les vecteurs u_j sont appelés* **vecteurs principaux** *de l'ACP.*

On déduit des 2 théorèmes la solution au problème de l'ACP. Pour tout $k \leq r$, le sev E_k engendré par les k premiers vecteurs principaux u_1,. . . , u_k est le sev principal de dimension k.

Ainsi :

— Une ACP *canonique ou simple* reposera sur la diagonalisation de la matrice de variance-covariance des p variables de départ ,

— Une ACP *normée*, reposera sur la diagonalisation de la matrice de corrélations des p variables de départ.

Propriété : On peut montrer que les inerties expliquées par ces axes sont égales aux valeurs propres λ_j :

$$I_{\Delta_{u_j}} = \lambda_j = I(\hat{\mathcal{N}}^{u_j}) .$$

D'après la décomposition de l'inertie expliquée, l'inertie expliquée par E_k est donc donnée par

$$I_{E_k} = \lambda_1 + \cdots + \lambda_k .$$

3 Analyse du nuage des individus

3.1 Composantes Principales (CP)

3.1.1 Définition

Rappelons que le point de départ était d'obtenir une représentation du nuage $\mathcal{N}$ dans des espaces de dimension réduite. On connait maintenant les axes définissant ces espaces. Pour pouvoir obtenir les différentes représentations, il suffit de déterminer les coordonnées des points du nuage dans la nouvelle base $M-$orthonormée constituée par les vecteurs principaux de l'ACP.

Pour tout $i = 1, \cdots, n$, soit c_i^j la coordonné sur l'axe Δ_{u_j} du projeté $M-$orthogonal de y_i sur Δ_{u_j}. La décomposition du vecteur y_i sur la base des vecteurs principaux $(u_1, u_2, \cdots, u_p)$ s'écrit alors

$$y_i = \sum_{j=1}^{r} c_i^j u_j .$$

On a classiquement :

$$c_i^j = < y_i, u_j >_M = {}^t y_i M u_j .$$

Définition 9. *Le vecteur de* $\mathbb{R}^n$

$$c^j = \begin{pmatrix} c_1^j \\ c_2^j \\ \vdots \\ c_n^j \end{pmatrix}$$

est appelé jème composante principale.

La projection du nuage $\mathcal{N}$ dans un "plan principal" $\Delta_{u_k} \oplus \Delta_{u_l}$ est obtenue par les composantes de c^k et de c^l càd
(c_i^k, c_i^l) = coordonnées sur les vecteurs u_k et u_l du projeté orthogonal de y_i dans le plan $\Delta_{u_k} \oplus \Delta_{u_l}$ (qu'on peut abréger en "plan principal (k,l)").

Définition 10. *Pour tout $k \leq p$ et $l \leq p$, $k \neq l$, la représentation des points $\{(c_i^k, c_i^l)\,;\, i = 1, \cdots, n\}$ dans le plan, avec c_i^k en abscisse et c_i^l en ordonnée, est appelé* **carte des individus sur le plan principal** (k, l).

Par exemple $\{(c_i^1, c_i^2)\,;\, i = 1, \cdots, n\}$ sont les coordonnées dans (u_1, u_2) du nuage projeté sur le 1er plan principal $\Delta_{u_1} \oplus \Delta_{u_2}$, dite carte des individus dans le plan $(1, 2)$.

Remarque : Pour deux individus h et i on a $d_M(y_h, y_i) = \|y_h - y_i\|_M = \sqrt{\sum_{j=1}^p (c_h^j - c_i^j)^2}$.

— Ainsi si y_h et y_i sont dans le plan principal (k, l) on a :

$$d_M(y_h, y_i) = \sqrt{(c_h^k - c_i^k)^2 + (c_h^l - c_i^l)^2}.$$

Donc utiliser la distance canonique du plan sur la carte des individus permet de calculer la distance $d_M(y_h, y_i)$.
Plus généralement le produit scalaire $<,>_M$ entre deux vecteurs du plan principal (k, l) revient au produit scalaire canonique de $\mathbb{R}^2$ entre les représentations de ces vecteurs sur la carte du plan principal (k, l).

— Si y_h et y_i ne sont pas dans le plan principal (k, l) on a :

$$d_M(y_h, y_i) \geq \sqrt{(c_h^k - c_i^k)^2 + (c_h^l - c_i^l)^2}.$$

Donc utiliser la distance canonique du plan sur la carte des individus donne une sous-estimation des distances $d_M(y_h, y_i)$, et la carte ne sera intéressante que si les individus sont "bien représentés" (voir plus loin).

3.1.2 Calcul des Composantes Principales

Proposition 3. *Les CP sont données par :*

Expression vectorielle : $c^j = YMu_j$

Expression matricielle : $C = YMU$,

où on a noté $C = [c^1, c^2, \cdots, c^p]$ *la matrice obtenue en rangeant en colonne les* c^j, *et* $U = [u_1, u_2, \cdots, u_p]$.

3.1.3 Composantes Principales : nouvelles variables

Une CP associe à chaque individu i un nombre réel. On peut donc la considérer comme une nouvelle variable. Comme les variables initiales y^j, cette variable est représentée par un vecteur de $\mathbb{R}^n$.

Proposition 4.
— *Les CP sont des combinaisons linéaires des variables de départ* $y^1, \ldots, y^p$.
— *Chaque CP* c^j *est centrées, de variance* λ_j
— *les CP sont non corrélées deux à deux.*

Remarque : Ainsi la matrice de variance-covariance des nouvelles variables c^j est diagonale.

3.1.4 Qualité de la représentation et contributions des individus

Rappelons que l'inertie totale du nuage $\mathcal{N}$ des individus vaut

$$I = \sum_{j=1}^{p} \lambda_j = tr(VM)\,.$$

Définition 11. *La qualité globale de la repésentation du nuage* $\mathcal{N}$ *sur le s.e principal* E_k *engendré par* $(u_1, \cdots, u_k)$ *est mesurée par le pourcentage d'inertie expliquée par* E_k

$$\frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{I} = \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_k}{\sum_{j=1}^p \lambda_j}\,.$$

Parallèlement à cet indice de qualité globale, on peut définir, pour chaque individu, la qualité de représentation sur un axe principal.

Définition 12. La qualité de représentation de l'individu i *sur l'axe principal* Δ_{u_k} *est mesurée par le cosinus carré de l'angle que fait* y_i *avec sa projection* $\hat{y}_i^{u_k}$ *sur* Δ_{u_k}

$$cos^2(y_i, \hat{y}_i^{u_k}) = \frac{||\hat{y}_i^{u_k}||_M^2}{||y_i||_M^2} = \frac{(c_i^k)^2}{\sum_{j=1}^p (c_i^j)^2}\,.$$

— Si $\cos^2(y_i, \hat{y}_i^{u_k})$ est proche de 1, l'individu i est bien représenté sur Δ_{u_k} .
— Si $\cos^2(y_i, \hat{y}_i^{u_k})$ est proche de 0, l'individu i est très mal représenté sur Δ_{u_k} .

Cette notion peut être généralisée en passant d'un axe à un sous-espace E_k. Par exemple la qualité de représentation de l'individu i sur le premier plan principal E_2 est mesurée par

$$\cos^2(y_i, \hat{y}_i^{E_2}) = \frac{||\hat{y}_i^{E_2}||_M^2}{||y_i||_M^2} = \frac{(c_i^1)^2 + (c_i^2)^2}{\sum_{j=1}^p (c_i^j)^2} = \cos^2(y_i, \hat{y}_i^{u_1}) + \cos^2(y_i, \hat{y}_i^{u_2})).$$

Remarque : Dans une carte des individus, on ne peut tirer des conclusions sur les individus (regrougements, individus exceptionels, etc...) que si ces individus sont bien représentés dans le plan principal considéré.

Contribution d'un individu à un axe : L'inertie globale portée par l'axe Δ_{u_k} vaut λ_k. Cette inertie se décompose de la manère suivante :

$$\lambda_k = Var(C^k) = \sum_{i=1}^n p_i (c_i^k)^2.$$

Ainsi :
$p_i(c_i^k)^2$ est la **contribution** de l'individu i à l'inertie portée par Δ_{u_k}
$\frac{p_i(c_i^k)^2}{\lambda_k}$ est la **contribution relative** de l'individu i à l'inertie portée par Δ_{u_k}.

On note que si tous les individus ont le même poids $\frac{1}{n}$ dans l'analyse, alors les contributions n'apportent pas plus d'information que les coordonnées.

Contribution relative d'un individu y_i à l'inertie du nuage :
On a par définition :

$$I = \sum_{i=1}^n p_i ||y_i||_M^2.$$

On définit ainsi la contribution d'un individu i à l'inertie totale du nuage :
$p_i \|y_i\|_M^2 = p_i \sum_{k=1}^n (c_i^k)^2$ est la **contribution** de l'individu i à l'inertie totale,
$\frac{p_i \|y_i\|_M^2}{I} = \frac{p_i \sum_{k=1}^n (c_i^k)^2}{I}$ est la **contribution relative** de l'individu i à l'inertie totale.

4 Analyse du nuage des variables

On s'intéresse dans cette section au nuage des variables défini par :

$$\mathcal{V} = \{y^1, y^2, \ldots, y^p\}.$$

Nous avons vu qu'il était intéressant d'utiliser la métrique des poids D pour calculer des distances entre éléments de $\mathcal{V}$, ou plus généralement entre des vecteurs centrés de $\mathbb{R}^n$. Il est important de noter

que l'analyse dans $\mathbb{R}^n$ ne se fait pas par rapport au centre de gravité du nuage points-variables mais par rapport à l'origine.

Rappelons que l'inertie totale du nuage des individus s'écrit :

$$I = tr({}^tYDYM).$$

On remarque aisément que c'est aussi :

$$I = tr(YM{}^tYD),$$

ce qui ressemble à la 1ère expression en inversant les rôles : les "individus" sont les variables, la "métrique" est D et la "pondération" est M.

On peut ainsi formuler un problème "d'ACP sur les variables" analogue au problème d'ACP sur les individus. On peut définir une inertie expliquée par un sev de $\mathbb{R}^n$, et l'objectif est de trouver les sous-espaces $F_1, F_2, \ldots$ de $\mathbb{R}^n$ qui maximisent cette inertie à dimension $1, 2, \ldots$ fixée. On les nommera espaces factoriels.

On passe ici sous silence la formulation et la résolution du problème "d'ACP sur les variables", et on se contente de donner les expressions des objets importants, qui sont tous en lien avec des objets déjà définis dans la Section 3.

4.1 Facteurs et axes, espaces (et "composantes") factoriels

Proposition 5. *1) la matrice $YM{}^tYD$ de taille n admet r valeurs propres positives non nulles*

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$$

et $n - r$ valeurs propres nulles (avec les λ_k et le r introduits dans la section précédente),

2) les r CP $(c^1, c^2, \ldots, c^r)$ correspondent à des vecteurs propres de la matrice $YM{}^tYD$ de taille n resp. associés aux valeurs propres $\lambda_1, \ldots, \lambda_r$.

On rappelle que les CP sont $D-$orthogonales, mais pas $D-$normées : On construit à partir des r CP non nulles, une base $D-$orthonormée du s.e.v engendré par les variables $(y^1, \cdots, y^p)$. Les vecteurs de cette base sont donnés par

$$\forall\, 0 \leq k \leq r\,, \quad v^k = \frac{1}{\sqrt{\lambda_k}} c^k\,.$$

Le vecteur v^k de $\mathbb{R}^n$ appelé **kème facteur**. Il est centré et normé.

On peut définir une inertie expliquée et appliquer la même démarche que celle utilisée dans l'espace des individus. On obtient les résultats suivants (admis) :

Proposition 6. *1) Pour tout $k \leq r$, le **s.e factoriel** F^k de dimension k de l'ACP sur les variables est le sev de $\mathbb{R}^n$ engendré par les k premiers facteurs $(v^1, v^2, \cdots, v^k)$.*

*2) L'inertie expliquée par l'axe engendré par le facteur v^k (nommé **axe factoriel** k) est λ_k, et donc celle expliquée par F^k est $\lambda_1 + \cdots + \lambda_k$*

Calcul des coordonnées des variables $y^1, y^2, \ldots, y^p$ sur le kème facteur :

Pour tout $j = 1, \cdots, p$, soit d_k^j la coordonné sur l'axe Δ_{v^k} du projeté $D-$orthogonal de y^j sur Δ_{v^k}. On a classiquement $d_k^j =< y^j, v^k >_D$.

La décomposition du vecteur y^j sur la base des facteurs $(v_1, v_2, \cdots, v_r)$ s'écrit alors

$$y^j = \sum_{k=1}^{r} d_k^j v^k\,.$$

On définit alors les "Composantes Principales de l'ACP sur les variables" (le terme "composante factorielle" serait légitime mais n'est pas utilisé) $d_1, d_2, \ldots, d_n$ comme $d_k = {}^t(d_k^1, d_k^2, \ldots, d_k^p)$.

On a une écriture vectorielle : $d_k = {}^t Y D v^k$,

et on démontre aussi un lien avec u_k :

$$d_k^j = < y^j, v^k >_D = \frac{1}{\sqrt{\lambda_j}} < y^j, c^k >_D = \frac{1}{\sqrt{\lambda_k}} < \sum_{l=1}^{p} u_l^j c^l, c^k >_D = \frac{1}{\sqrt{\lambda_k}} \sum_{l=1}^{p} u_l^j < c^l, c^k >_D .$$

Les c^k étant centrées, non corrélées et de variance λ_k, on obtient

$$d_k^j = \sqrt{\lambda_k} u_k^j , \text{ pour tout } k = 1, \ldots r \text{ et } j = 1, \ldots, p .$$

On a donc

$$d_k = \sqrt{\lambda_k} u_k , \qquad \text{pour tout } k = 1, \ldots r$$

et d_k est égal au vecteur nul de $\mathbb{R}^p$ pour $k = r+1, \ldots, n$.

Remarque : les vecteurs $d_1, d_2, \cdots, d_r$ sont des vecteurs propres de VM.

Définition 13. *Pour tout $k \leq r$ et $l \leq r$, $k \neq l$, la représentation des points $\{(d_k^j, d_l^j)\,;\, j = 1, \cdots, p\}$ dans le plan, avec d_k^j en abscisse et d_l^j en ordonnée, est appelé* **carte des variables sur le plan factoriel** (k, l).

Remarque 1 : Dans une carte des variables, on choisit souvent de représenter les variables par des vecteurs partant de l'origine. Cela permet de visualiser les normes, produits scalaires et angles des variables projetées, qui jouent un rôle important.

Remarque 2 : Pour deux variables j et j' on a $< y^j, y^{j'} >_D = \sum_{k=1}^{n} d_k^j d_k^{j'}$. Ainsi si y^j et $y^{j'}$ sont dans le plan factoriel (k, l) on a :

$$< y^j, y^{j'} >_D = d_k^j d_k^{j'} + d_l^j d_l^{j'} = < (d_k^j, d_l^j), (d_k^{j'}, d_l^{j'}) >,$$

où $<,>$ est le produit scalaire canonique de $\mathbb{R}^2$.

Ainsi les produits scalaires (et donc aussi les normes et angles) des variables initiales sont égales aux produits scalaires (et normes et angles) calculés sur la carte (k, l) avec la géométrie classique du plan.

Par contre cela devient faux pour des variables mal représentées, notamment les distances vues sur la carte sous estiment les vraies distances entre variables initiales.

4.2 Qualité de la représentation (et contributions) des variables

Définition 14.

— **La qualité globale** *de la représentation du nuage $\mathcal{V}$ sur le s.e factoriel F^k est mesurée par* $\dfrac{\lambda_1 + \cdots + \lambda_k}{\sum_{j=1}^{p} \lambda_j}$.

— **La qualité de représentation de la variable y^j sur l'axe factoriel engendré par v^k est mesurée par**

$$\cos^2(y^j, \hat{y}^{j,v^k}) = \frac{||\hat{y}^{j,v^k}||_D^2}{||y^j||_D^2} = \frac{< y^j, v^k >_D^2}{s_j^2} = r^2(y^j, c^k) ,$$

où $r^2(y^j, c^k)$ est le coefficient de corrélation linéaire entre y^j et c^k.

Exemple : La qualité de représentation de la variable y^j sur le premier plan factoriel F_2 engendré par v^1 et v^2 est mesurée par :

$$\cos^2(y^j, \hat{y}^{j,F_2}) = \frac{||\hat{y}^{j,F_2}||_D^2}{||y^j||_D^2} = r^2(y^j, c^1) + r^2(y^j, c^2)\,.$$

— Si $\cos^2(y^j, \hat{y}^{j,F_2})$ est proche de 1 alors la variable y^j est bien représentée dans F_2.
— Si $\cos^2(y^j, \hat{y}^{j,F_2})$ est proche de 0 alors la variable y^j est très mal représentée sur F_2.

Remarques sur la notion de contribution d'une variable à une inertie

On a vu que $d_k = \sqrt{\lambda_k}u_k$, avec u_k $M-$normé, et donc $||d_k||_M^2 = \lambda_k$. De plus on a vu que $I = tr(YM{}^tYD)$. Toutefois, contrairement au cas des individus, ces relations ne permettent pas de définir des contributions des variables à un axe ou à l'inertie totale.

Cependant, dans le cas où la métrique est diagonale $M = Diag(m_1, \ldots, m_p)$, ces relations conduisent à :

$$\lambda_k = \sum_{j=1}^{p} m_j (d_k^j)^2 \quad \text{et} \quad I = \sum_{j=1}^{p} m_j ||y^j||_D^2.$$

Ainsi :

$m_j(d_k^j)^2$ joue le rôle de la contribution de la variable j à l'inertie portée par l'axe factoriel k,

$\frac{m_j(d_k^j)^2}{\lambda_k}$ joue le rôle de la contribution relative de la variable j à l'inertie portée par l'axe factoriel k,

$m_j||y^j||_D^2 = m_j var(y^j)$ joue le rôle de la contribution de la variable j à l'inertie totale,

$\frac{m_j||y^j||_D^2}{I}$ joue le rôle de la contribution relative de la variable j à l'inertie totale.

4.3 Particularités de l'analyse des variables en ACP normée

On se place ici dans le cadre d'une ACP normée. Rappelons que ceci revient à centrer et réduire les données en colonne au préalable de l'analyse et prendre pour métrique $M = I_p$ dans l'espace des individus. Considérons Z le tableau centré réduit. On a

$$Z = [z^1, z^2, \ldots, z^p]\,,$$

où pour tout $j = 1, \ldots, p$, $\quad z^j = \frac{x^j - \bar{x}^j}{s_j}$.

4.3.1 CP de l'ACP sur les variables

Soit d_k^j la coordonnée de z^j sur le kème axe. Etant donné que les vecteurs z^j et v^k sont normés on a :

$$d_k^j = < z^j, v^k >_D = r(z^j, v^k)\,.$$

Remarque : On a $v^k = \frac{1}{\sqrt{\lambda_k}}c^k$ et la corrélation est insensible aux facteurs multiplicatifs. Donc on a aussi :

$$d_k^j = r(z^j, c^k)\,.$$

Ainsi **en ACP normée d_k^j est la corrélation entre la variable initiale z^j et la nouvelle variable synthétique c^k.**

On montre que les quantités d_k^j jouent un rôle majeur dans l'analyse des variables en ACP normée :

d_k^j est la coordonnée de z^j sur le kème axe factoriel,

d_k^j est la corrélation entre la variable initiale z^j et la nouvelle variable synthétique c^k,

$(d_k^j)^2$ est la qualité de représentation de la variable z^j sur le kème axe factoriel,

$(d_k^j)^2$ (respectivement $\frac{(d_k^j)^2}{\lambda_k}$) peut être considéré comme la contribution (resp contribution relative) de la variable j à l'inertie portée par le kème axe factoriel.

4.3.2 Carte des variables : cercles des corrélations

On s'intéresse à la carte des variables obtenue par la projection du nuage des variables $\mathcal{V}$ sur le plan factoriel engendré par v^k et v^l. La projection de z^j sur ce plan sera notée ici $\hat{z}^j$.

Dans une ACP normée on peut voir que :

$$d_D(0, z^j) = ||z^j||_D = 1.$$

Ainsi dans l'espace $\mathbb{R}^n$, **les variables centrées réduites sont toutes situées sur une hypersphère de centre** 0 **et de rayon** 1 appelée **hypershère des corrélations**. Les plans principaux couperont cette hypersphère suivant de grands cercles de rayon 1, **les cercles des corrélations**, à l'intérieur desquels se trouveront les $\hat{z}^j$, projections des variables sur ces plans.

Proposition 7. *Etant donné que* $||z^j||_D^2 = 1$*, la qualité de représentation de la variable* z^j *sur le plan factoriel* (k, l) *est mesurée par la quantité :*

$$\cos^2(\hat{z}^j, z^j) = ||\hat{z}^j||_D^2 = r^2(z^j, c^k) + r^2(z^j, c^l)\,.$$

Une variable est donc bien représentée sur le plan factoriel (k, l) *si sa projection est* **proche du cercle des corrélations**.

Comme cela a déjà été mentionné, on peut utiliser la géométrie classique de $\mathbb{R}^2$ sur une carte des variables, pour étudier les variables initiales (du moment qu'elles sont bien représentées). Par géométrie classique on entend :

$$\forall a = {}^t(a_1, a_2) \in \mathbb{R}^2, \quad \forall b = {}^t(b_1, b_2) \in \mathbb{R}^2,$$
$$< a, b >= a_1 b_1 + a_2 b_2, \qquad \text{et} \qquad d(a, b) = \|a - b\| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}.$$

En particulier, les cosinus des angles de la carte reflètent les cosinus de l'espace des variables, qui sont aussi les corrélations. Ainsi, si deux variables z^j et $z^{j'}$ sont bien représentées par leurs projections $\hat{z}^j$ et $\hat{z}^{j'}$ sur le plan factoriel (k, l) alors :

- la proximité des projections $\hat{z}^j$ et $\hat{z}^{j'}$ indique une forte corrélation linaire entre z^j et $z^{j'}$,
- des points $\hat{z}^j$ et $\hat{z}^{j'}$ diamétralement opposés indiquent une corrélation négative proche de -1,
- des directions de $\hat{z}^j$ et de $\hat{z}^{j'}$ presque orthogonales indiquent une faible corrélation entre z^j et $z^{j'}$.

La carte des variables sur le plan (k, l) donne aussi des informations sur les liens entre les variables initiales z^j et les nouvelles variables c^k et c^l. Ces liens permettent d'interpréter les nouvelles variables en termes de corrélations avec les variables initiales.

- Si sur la carte, $\hat{z}^j$ est proche du point $(1, 0)$ (respectivement $(-1, 0)$), alors z^j est très corrélée positivement (resp négativement) avec c^k.
- Si sur la carte, $\hat{z}^j$ est proche du point $(0, 1)$ (respectivement $(0, -1)$) alors z^j est très corrélée positivement (resp négativement) avec c^l.
- Si sur la carte, $\hat{z}^j$ est proche de l'origine, alors z^j est peu corrélée avec c^k et c^l.

4.3.3 Liens entre carte des individus et carte des variables

Certaines positions sur la carte des variables donnent des informations sur le nuage des individus centrés réduits, notamment en ce qui concerne les positions des vecteurs principaux par rapport aux axes canoniques de $\mathbb{R}^n$. Pour j donné, soit Δ_j le jème axe de $\mathbb{R}^n$ (celui où l'on marque les valeurs $z_1^j, \ldots, z_n^j$ pour former le nuage des individus).

Proposition 8. — *Si sur la carte, $\hat{z}^j$ est sur le point $(1,0)$: le vecteur u_k appartient à Δ_j (si du moins $r = p$).*
Si plus généralement, sur la carte, $\hat{z}^j$ est sur le cercle des corrélations, alors $\Delta_j \subset \Delta_{u_k} \oplus \Delta_{u_l}$. Alors pour tout $i \in \{1, \ldots, n\}$, $\hat{z}_i^j = z_i^j$: la jème variable est parfaitement représentée dans le nuage projeté sur le plan principal (k,l) (si du moins $r = p$).

— *Si sur la carte, $\hat{z}^j$ est à l'origine $(0,0)$, alors Δ_j est orthogonal à u_k et u_l. Alors pour tout $i \in \{1, \ldots, n\}$, $\hat{z}^j = 0$: la jème variable est constante (nulle) dans le nuage projeté sur le plan principal (k,l).*

On peut par ailleurs faire des liens entre la carte des variables sur (k,l) et la carte des individus sur (k,l).

Les individus situés le plus à droite sur la carte des individus ont des valeurs c_i^k fortement positives. D'après le cours sur la corrélation, ils auront tendance à avoir :

— des valeurs fortement positives pour les variables initiales à droite de la carte des variables (proches du point $(1,0)$)

— des valeurs fortement négatives pour les variables initiales à gauche de la carte des variables (proches du point $(-1,0)$)

et inversement pour les individus les plus à gauches sur la carte des individus.

Les individus situés le plus en haut sur la carte des individus ont des valeurs c_i^l fortement positives. Ils auront tendance à avoir :

— des valeurs fortement positives pour les variables initiales en haut de la carte des variables (proches du point $(0,1)$)

— des valeurs fortement négatives pour les variables initiales en bas de la carte des variables (proches du point $(0,-1)$)

et inversement pour les individus les plus en bas sur la carte des individus.

Ces constatations permettent de donner du sens aux nouvelles variables, en termes d'opposition entre des groupes d'individus.

5 Pratique de l'ACP

5.1 Nombre d'axes à retenir

Le principal objectif d'une ACP étant la réduction du nombre de variables initiales, la détermination du nombre k d'axes à retenir est donc très importante. De nombreux critères de choix pour k ont été proposés dans la littérature. Voici les plus courants.

5.1.1 Part d'inertie

Souvent la qualité globale des réprésentations est utilisée pour choisir k de sorte que la part expliquée soit supérieure à une valeur seuil fixée a priori par l'utilisateur.

5.1.2 Règle de Kaiser

Elle préconise de conserver que les valeurs propres supérieures à leur moyenne I/p car seules jugées plus "informatives" que les variables initiales. Dans le cas d'une ACP normée, ne sont retenues que les valeurs propres supérieures à 1. Ce critère utilisé implicitement par **SAS/ASSIST**, a tendance à surestimer le nombre de composantes pertinentes.

5.1.3 Eboulis des valeurs propres

C'est un graphique présentant la décroissance des valeurs propres obtenu en traçant les valeurs propres λ_j en fonction de leur indice j. Le principe consiste à chercher, s'il existe un "coude" dans le graphe. Les axes à retenir sont alors ceux dont les valeurs propres se situent avant le "coude".

5.2 Interprétation

Définition 15. *On appelle* **élément supplémentaire** *un élément (individu ou variable) qui n'a pas été pris en compte dans l'analyse pour la détermination des axes (dans le cas d'une ACP normée, on calcule la matrice de corrélation sans lui), mais dont on a calculé ensuite ses composantes sur chacun des axes pour le porter sur les graphiques.*

Pour décrire une carte des variables ou des individus, on adoptera le plan suivant :

1- Donner le pourcentage d'inertie expliquée par le plan considéré et chacun des axes,

2- Indiquer les variables (resp.les individus) mal représentées dans ce plan pour les exclure de la description,

3- Utiliser les *contributions*
   - *des variables* (si ces contributions sont bien définies) pour interpréter les axes en termes de variables de départ. Pour chaque axe factoriel important, on pourra lister les variables initiales de plus forte contribution à cet axe, et les répartir dans un tableau à deux entrées selon le signe (positif ou négatif) de leur coordonnée sur cet axe.
   - *des individus* pour identifier ceux qui sont influents pour l'orientation d'un axe et ceux qui ont une contribution *excessive* qui notamment pourraient être un facteur d'instabilité (le fait d'enlever un tel individu de l'analyse modifiant de manière importante les résultats). Il est important de vérifier qu'il ne s'agit pas de données éronnées et de faire une nouvelle analyse en les considérant en *supplémentaires*.

4-a **Pour une carte des variables** : étudier les angles entre les projections des variables en termes de covariance ou de corrélation (selon le type d'ACP choisi) pour dégager éventuellement des *groupes* de variables.
Vérifier les tendances visualisées sur la carte par un examen de la matrice de corrélation (pour l'ACP normée).

4-b **Pour une carte d'individus** : étudier les proximités ou les oppositions entre les points en termes de "comportement" et dégager éventuellement des *groupes* d'individus et des comportements singuliers de certains. Vérifier les caractéristiques dégagées par un examen des données de départ.

5- Faire une synthèse des informations et hypothèses principales dégagées de la carte décrite.

5.3 Effet de "taille"

Théorème 3 (Frobenius)**.** *Une matrice symétrique dont tous les termes positifs admet un premier vecteur propre dont toutes les composantes sont de même signe.*

Application à l'ACP simple ou normée : Si les variables sont toutes corrélées positives entre elles, la matrice de variance-covariance et la matrice de corrélation ont tous leurs termes positifs. Donc d'après le théorème de Frobenius, la première composante principale est corrélée positivement avec toutes les variables (si on choisit le signe positif pour les composantes du premier vecteur propre) et les individus sont rangés sur le premier axe principal par valeurs croissantes de l'ensemble des variables (en moyenne).

Ainsi lorsque toutes les variables x^j sont corrélées positivement entre elles, on dit que la première composante principale c^1 définit un facteur de "taille".

La deuxième composante principale c^2 différencie alors les individus de "taille" similaires. Il est appelé **facteur de "forme"**.